
Like

STATUS UPDATE: STRESSED, ANGRY, AT RISK?

Using powerful new tools, scientists are mining social media to assess mental and physical health from afar

By Johannes C. Eichstaedt

Late one night in 2010 after the annual Google Zeitgeist conference had wrapped up, psychologist Martin Seligman, a featured speaker that year, found himself in a huddle with some of the biggest names in technology. Google had just broken ground using search-engine queries to monitor the spread of influenza in the U.S., and Google Maps was taking the world by storm. The potential applications for such tools seemed limitless, so Seligman, a founding father of positive psychology, and Google co-founder Larry Page, among others, began to explore the possibilities.

Illustration by **PAUL OAKLEY**

What if something like Google Flu Trends could be developed to chart psychological health in America? Specifically, they wondered if a computer algorithm could accurately predict happiness and well-being across time and space by tracking the language that people used on social media.

In a matter of months Seligman, along with me and a few others at the University of Pennsylvania, launched the World Well-Being Project (WWBP), initially in collaboration with Google.org, the company's philanthropic arm. Since then our cross-disciplinary team of psychologists, computer scientists, statisticians and app developers has grown rapidly. It now includes 13 staff scientists pursuing 45 subprojects with collaborators in governments and organizations around the world.

In January 2015 my colleagues and I published an important proof-of-concept result. We evaluated more than 100 million tweets from more than 1,300 counties across the U.S., a sample of public data available from Twitter. We discovered that the preponderance of negative tweets—particularly those expressing anger or hostility—in a given location reliably predicted rates of death from heart disease there. Many other findings have piled up that reveal associations between the language in tweets or Facebook posts and traits that include age, gender, personality and income level, as well as mental illness and physical ailments.

The language people use in tweets and Facebook posts can indicate their age, gender, personality, income, mental state and physical health.

These results make it clear that social media data are a rich resource that psychologists, sociologists, epidemiologists and others can mine to make valuable communitywide health forecasts and even individual diagnoses. The opportunity may be huge, but this fast-growing field has a dark side, too. Analyses of how people use language on social media are based solely on statistical patterns. But they can be so revealing that intelligence agents, political candidates and businesspeople—from marketers to insurance actuaries—are just as interested as scientists in their application. Indeed, few people realize just how much information algorithms can cull from their routine activity on Facebook and Twitter.

FAST FACTS

USING SOCIAL MEDIA TO MAKE DIAGNOSES

1. Social media feeds offer psychologists, sociologists and epidemiologists, among others, a new resource for tracking health and well-being at a distance.
2. Algorithms that make use of machine learning can analyze the language people use online to craft detailed psychological profiles of individuals and entire communities.
3. Using these methods, researchers are finding a wide range of revealing associations between social media posts and personality, mental health, and even physical conditions such as heart disease.

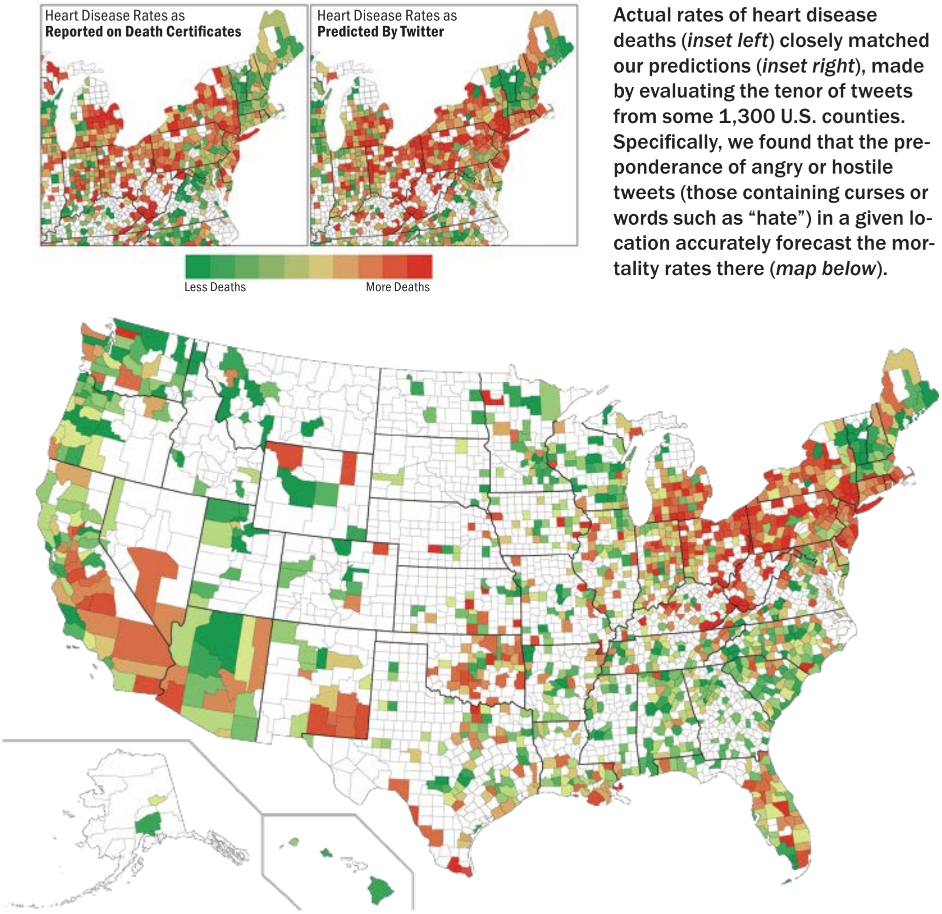
What Words Can Reveal

Before the WWBP team began testing tweets to spot health trends, Google had taken an intriguing first step. In 2008 physician Roni Zeiger, then the firm's chief health strategist, and his co-workers launched Google Flu Trends. The program spotted Google searches on terms related to flu symptoms and remedies and noted where geographically those searches were entered. In this way, they could plot the spread of infection in real time. Remarkably, their tracking of the flu season matched the statistics collected by the Centers for Disease Control and Prevention—only Google had the information first because they did not need to wait for doctors and hospitals to report each outbreak. Google discontinued Flu Trends last year, but the program demonstrated that search queries offered a viable way to monitor the spread of specific conditions, and it kick-started the field now called digital epidemiology. Newer endeavors are exploring search queries as a means to surveil not just influenza but dengue fever, malaria and sexually transmitted diseases.

Understanding the psychological states of entire populations, as our project strives to do, can be a little more nuanced than tracking illness: no one Googles "I am happy" in the way they search for remedies when they feel unwell. So we have had to take a less direct route, analyzing what people write on social media instead of their search terms. Decades of research have found that the words people choose in everyday conversation can reveal a great deal about their underlying psychology. And there are countless links between a person's mental state and physical well-being. Stress, negativity, anxiety and depression can impair our immune and cardiovascular systems, for example. Likewise, positive emotions and optimism appear to have a protective effect, reducing the risk of many diseases, including atherosclerotic heart disease, and increasing life expectancy.

Beginning in the 1990s, social psychologist James Pennebaker, now Regents Centennial Professor at the University of Texas at Austin, and his colleagues made a series of intriguing discoveries about the link between words and well-being. They were investigating why people who write about a trauma after the fact—a technique called expressive writing—are less likely to become physically ill, compared with people who keep disturbing experiences secret. To evaluate what their subjects penned, they used a computer program to quickly tally the words and concepts these essays contained.

Much to their surprise, they found that the actual content of the writing—whether it included positive or negative language and ideas—uncovered less about a participant's mental health than functional parts of speech did. For example, they



Actual rates of heart disease deaths (*inset left*) closely matched our predictions (*inset right*), made by evaluating the tenor of tweets from some 1,300 U.S. counties. Specifically, we found that the preponderance of angry or hostile tweets (those containing curses or words such as "hate") in a given location accurately forecast the mortality rates there (*map below*).

COURTESY OF JOHANNES C. EICHSTAEDT

found that people in the throes of depression did not necessarily write about sad things, but they did reliably use more first-person-singular pronouns: I, me, mine. Depressed patients tend to ruminate and are often heavily preoccupied with their own suffering. They found that the symptoms of trauma frequently lifted when participants started using more causal words (because, therefore, but) and complex language in their writing. These patterns appeared to signal that a patient was starting to make sense of the trauma and integrate it into a coherent narrative.

As Facebook, Twitter and similar applications have taken off during the past decade, the amount of language data available for analysis has expanded dramatically, offering psychologists a vast new window into the mental health of social media users. (In general, we work with anonymized data and request permission from Facebook users.) Of course, people present a version of themselves online, playing up good behaviors, traits and events—a skew that researchers refer to as a social desirability bias. But such biases are often less distorting than you might think. The fact that people are Facebook friends with their real friends usually helps to keep them fairly honest online. Real-world acquaintances know that their life is not all picnics and parties. Moreover, because these biases tend to affect everyone equally, algorithms can still tease out key differences among people. In support of this idea, when we compare the predictions from our methods against data collected in traditional ways (phone surveys, hospital reports, and so on), which sample the population more representatively, we have often been surprised just how close the two are.

Needles in a Haystack of Tweets

The language on social media feeds—filled with slang and emoticons—presents a challenging set of data to parse. We are only able to extract meaningful patterns from it today thanks to breakthroughs in natural-language processing during the past 30 years. With exponential growth in computing power, it has become increasingly possible to process language using statistical pattern-recognition algorithms, also called machine learning. These tools have matured dramatically in recent years and enable applications such as Apple's Siri and Google's analytics programs, which drop those eerily appropriate ads alongside your e-mail's in-box.

Before these advances, language analysis in psychology was based on simpler dictionary-based approaches that linked emotional states to predetermined lists of words. For instance, if the word "happy" appeared in a text, it was taken as a sign of positive emotion. This method sometimes yielded confusing results because language is inherently ambiguous. When early work in this area applied a psychologist-made "positive emotion" dictionary to social media feeds, it erroneously indicated that there were huge spikes in happiness on New Year's Day—simply because people were wishing one another a "happy" new year.

THE AUTHOR

JOHANNES C. EICHSTAEDT is a data scientist pursuing a Ph.D. in psychology at the University of Pennsylvania, where he co-founded the World Well-Being Project in 2011.

How Twitter Language Reflects the Risk of Heart Disease

In an analysis of public tweets by U.S. county, we found that words associated with negative topics (*shown in red on left*) were correlated with higher rates of death from heart disease. More optimistic subjects (*shown in green on right*) were more often tweeted from communities with lower rates. We found the strongest language correlations with atherosclerosis, the leading cause of death in the U.S., often thought to be closely linked to psychological factors.

Modern machine-learning approaches avoid these errors. They start out agnostically—that is, they make no assumptions whatsoever about which words indicate what emotions or traits—and then they cluster, count, score and isolate words to "learn" psychological associations from scratch. One shortcoming is that these methods work only on data sets with at least 5,000 to 10,000 users. Indeed, the more entries, the more accurate the results because additional input allows us to isolate even faint signals amid all the noise of daily posts. Fortunately for us, social media sites now have hundreds of millions of users.

In 2013 H. Andrew Schwartz, now at Stony Brook University, Seligman and I, along with our colleagues, published a study in which we had applied a machine-learning method to 700 million words, phrases and topics gleaned from the Facebook messages of 75,000 volunteers, who also took personality tests. To date, it was the largest study linking language and personality, by an order of magnitude. Once the algorithms had the status updates of so many Facebook users and knew how these users scored on a personality test, they could correlate words to personality traits.

Using the results, we created word clouds to show the words that best distinguished extroversion from introversion and neuroticism from emotional stability. We found that some words are rarely used, but when they are, they are highly predictive of a psychological trait. For example, the use of "depressed" is a powerful, if seldom seen, marker of neuroticism.

Many of the associations made sense, but some were surprising. You may have predicted that extroverts more often use the word "party" and introverts "computer," but would you have guessed that the word "apparently" is used far more often by people who score high in neuroticism? Or that people who are emotionally stable write more frequently about sports? Or that introverts show a greater interest in Japanese media, such as anime, and emoticons? In a follow-up study led by our colleague Gregory Park, we applied the algorithms to another set of Facebook posts to actually predict users' personalities using only their feeds. Remarkably, the algorithms did as well as or better than friends who filled out personality surveys about the subjects. In a sense, the algorithms wound up knowing these people better than their friends did! We took this result as a good sign that we had a handle on desirability biases.

My colleagues and I then used this same approach on Twitter to estimate an average "psychological profile" for every U.S. county, without having to knock on anyone's door [*see box at left*]. The counties whose tweets expressed more negative emotion, anger and hostility—filled with words such as "hate" and curse words—had the most heart disease deaths, too, according to CDC data on causes of death, based on actual death certificates. Optimistic counties had lower heart disease–related mortality rates. When we drilled down further into the data, we realized that our method had worked particularly well at predicting deaths from atherosclerosis, more so than for other forms of heart disease. Atherosclerosis is the leading cause of death in the U.S. and, not surprisingly, the kind of cardiac disease thought to be most associated with psychological causes.

Curiously, the people tweeting were not the people dying. Our method said nothing about anyone's individual risk for

COURTESY OF JOHANNES C. EICHSTAEDT

atherosclerosis. Rather the overall tenor of the tweets—typically from people too young to have heart problems—appear to have captured a snapshot of the psychology of the community at large. It is possible that these tweets detected an elusive quality of communities sociologists often refer to as social cohesion, loosely defined as the willingness of community members to cooperate and help one another. Research indicates that the more cohesive a community, the more its members tend to enjoy good mental and physical health. At present, we are extending our analysis to look for words and language patterns predictive of all the leading causes of death, including cancer and stroke. But we expect that here, too, more positive—and presumably more cohesive—communities will fare better.

Using only the language on Facebook feeds, researchers can predict people's personalities—sometimes better than their friends can.

Creating Health Dashboards

Our Twitter-based prediction of community-wide deaths from heart disease proved more accurate than any made using government statistics for known risk factors, including obesity, diabetes, smoking and hypertension. The result was so robust, in fact, that our language variables could predict rates of heart disease even after we controlled for strong classical predictors, such as education and poverty.

As more people make use of social media, our predictions may become even better. Ten years ago only certain segments of the population were Facebook users—mostly teens and young adults. As of October 2015, though, the Pew Research Center reported that 65 percent of American adults use social media sites regularly—a 10-fold increase since 2005. Fully 90 percent of younger adults use social media, and rates among those older than 65 have more than tripled since 2010. The median age of Twitter users is 32—only six years younger than the median age of the U.S. population.

Compared with the CDC's Behavioral Risk Factor Surveillance System or Gallup polls—which take into account much smaller samples at far greater cost—Twitter- or Facebook-based health assessments may offer a faster, cheaper dashboard indicator of community-wide well-being. My colleagues and I have now consulted with colleges in the U.K., the state government of South Australia and Mexican authorities who, with aspirations of tracking health trends via Twitter, may begin asking for people's social media handles in their next national census. It is, in many ways, a natural step in a world in which increasingly fewer people are reachable to survey via landlines.

Analyses of language on social media might also be applied to help clinicians treat individual patients. Lyle Ungar and others from our project recently teamed up with Raina Merchant and other Penn Medicine colleagues to put iPads in the Hospital of the University of Pennsylvania's emergency room. We asked ER patients to voluntarily sign in to Facebook and give permission for their status updates to be analyzed. Using our machine-learning methods, we then correlated all the language patterns with their medical records. Looking at the data, we have discovered an array of potential language markers for various diseases—including depression. In the future, doctors might be able to analyze social media posts for a disease's linguistic red flags and follow up with patients appropriately.

Indeed, it seems entirely conceivable that phone apps analyzing the narrative of our lives could help physicians deliver better care, across different contexts. Imagine a psychotherapist who receives an automatic, daily mood reading from his or her depressed patients and can then text them a critical reminder, insight or urgent recommendation. Or a doctor monitoring social media feeds for signs of depression among heart attack patients, for whom depression is a big risk factor for having a repeat crisis. In 2013 a team at Microsoft Research used tweets to predict postpartum depression among 376 new mothers. Their model was 71 percent accurate when it analyzed only prenatal posts and about 80 percent accurate when it also included tweets from the first few weeks following birth.

These are far better uses than, say, an insurance provider or loan shark surreptitiously using social media analytics to deny services or increase rates. In our research, we always obtain permission to analyze participants' online feeds and follow strict privacy guidelines. But few Facebook users realize that giving access to their statuses—or even just their "likes"—can supply a corporation with a fairly fine-grained personality profile. Societies have a way of co-evolving along the lines of their most powerful technologies—and it will take many of us across science, policy and industry to get this one right. The more we realize the potential of social media analytics to better our health and well-being, though, the more we can join in efforts to craft our future in ways that are conscious, ethical and even lifesaving. M

MORE TO EXPLORE

- **Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach.** H. Andrew Schwartz et al. in *PLOS ONE,* Vol. 8, No. 9, Article No. e73791; September 25, 2013.
- **Psychological Language on Twitter Predicts County-Level Heart Disease Mortality.** Johannes C. Eichstaedt et al. in *Psychological Science,* Vol. 26, No. 2, pages 159–169; February 2015.
- **Linking Social Media and Medical Record Data: A Study of Adults Presenting to an Academic, Urban Emergency Department.** Kevin A. Padrez et al. in *BMJ Quality & Safety.* Published online October 13, 2015.

From Our Archives

- **A Digital Safety Net.** Roni Jacobson; November/December 2014.